Credibility: How Agents Can Handle Unfair Third-party Testimonies in Computational Trust Models

Jianshu Weng, Zhiqi Shen, Chunyan Miao, Angela Goh, and Cyril Leung

Abstract—Usually, agents within multi-agent systems represent different stakeholders that have their own distinct and sometimes conflicting interests and objectives. They would behave in such a way so as to achieve their own objectives, even at the cost of others. Therefore, there are risks in interacting with other agents. A number of computational trust models have been proposed to manage such risk. However, the performance of most computational trust models that rely on third-party recommendations as part of the mechanism to derive trust is easily deteriorated by the presence of unfair testimonies. There have been several attempts to combat the influence of unfair testimonies. Nevertheless, they are either not readily applicable since they require additional information which is not available in realistic settings, or ad-hoc as they are tightly coupled with specific trust models. Against this background, a general credibility model is proposed in this paper. Empirical studies have shown that the proposed credibility model is more effective than related work in mitigating the adverse influence of unfair testimonies.

Index Terms—Agent; Trust; Credibility; Unfair Testimonies

1 INTRODUCTION

Recent years have seen an increasing number of agents being developed to extend the sphere of human interaction within environments such as the Internet [1]. Those agents, with their autonomous reasoning and decision-making capability, can engage in complex interactions on behalf of their owners.

There is no single-agent system [2]. Instead, agents usually live in a society of agents, which is known as multi-agent system (MAS) [3], [4]. Usually, agents in a MAS represent various stakeholders, each with distinct interests and objectives. They try to pursue their own objectives, even at the cost of others. Hence, there are risks when agents interact with one another. Given this, trust, which is essential in regulating interactions in human society, needs to be established to regulate the interactions within MAS [5], [6], [7].

1.1 Computational Trust

Trust has long been a research topic in various disciplines such as sociology, economics, and philosophy (see [8] for a comprehensive review). The growth of multi-agent system (MAS) and its close resemblance to human societies have made trust an important topic in MAS too. The primary focus of research on trust in MAS is to convert trust into a computational concept so that agents are able to quantify the amount of trust they should place in others. Trust, as a computational concept, can be illustrated by a relational model as shown in Fig. 1.

Fig. 1: Graphical representation of trust, from the truster’s perspective

Fig. 1 captures the common understandings of computational trust. First of all, it involves two agents: truster and trustee. The trustee is an agent whose trustworthiness is being evaluated, while the truster is the one who is evaluating the trustee’s trustworthiness1. Secondly, trust is an attribute associated with interaction. It is about the truster’s decision to interact with the trustee given the uncertainty regarding the trustee’s behavior [9]. The truster’s evaluation of the trustee’s trustworthiness quantifies such uncertainty as its subjective evaluation of the probability that the latter will behave cooperatively in the interaction. Such an evaluation is based on the trustee’s past behavior, which is an attribute of

1. The truster is not visible in Fig. 1 since it is an illustration from the truster’s perspective.
the trustee. If the truster decides to interact with the trustee, the interaction will be associated with another attribute, i.e., the outcome, which captures the trustee’s real behavior in that particular interaction. This outcome will also be used to update the trustee’s attribute, which will be used to evaluate its trustworthiness for future interactions. Thirdly, trust is context-sensitive. Trust is associate with interaction, which can occur in various contexts. The trustee can have different behaviors in different contexts. Therefore, the truster’s evaluation of the trustee’s trustworthiness in one context is not necessarily the same as that in another. For example, truster $a_o$ trusts $a_o$ in repairing a car, though it does not trust $a_o$ at all in fixing computer hardware problem.

1.2 The issue of unfair testimonies

Based on the above common understandings, a number of computational trust models have been proposed. In these models, the information about the trustee’s “past behavior” typically comes from two sources: (1) the trustee’s previous personal interaction experience with the truster, and (2) third-party testimonies on the corresponding trustee. The former is the most direct source. Nevertheless, there are cases where the truster does not have any previous interaction experience with the trustee, or the past interaction experience is insufficient to support the trust evaluation. In such cases, third-party testimonies are sought by the truster.

A problem arises when third-party testimonies are aggregated due to the potential presence of unfair testimonies, i.e., those that do not truly reflect the trustee’s real behavior as perceived by the testimony receiver (i.e. the truster). An unfair testimony can occur due to the different views between the witness (i.e. the agent who gives testimony) and the truster. It can also result from the differing or even conflicting interests of the two agents. In both cases, unfair testimonies, if used blindly, will bias a truster’s trust evaluations [7]. Hence, the avoidance or mitigation of the influence of unfair testimonies has been identified as one of the fundamental issues in the research on computational trust [7].

1.3 The proposed credibility model

To address the issue of unfair testimonies, this paper proposes a model that helps the trusters to evaluate the credibility of the witnesses who provide testimonies on the trustees. Furthermore, the proposed model also guides the trusters how to filter and aggregate testimonies based on the evaluation of credibility.

As discussed earlier, the truster’s evaluation of the trustee’s trustworthiness quantifies its uncertainty about the trustee’s behavior in future interactions. Here, credibility measures how useful the witness’ testimony is in reducing such uncertainty. The relationship between these two concepts can be illustrated in Fig. 2. From this figure, it is observed that credibility is decoupled from trust. It is also observed that, besides the common agent properties, the witness also has an attribute past testimony-reporting. Information associated with this attribute is the only information needed to evaluate the corresponding witness’ credibility. In contrast, most of the related work has limited applicability since they usually require additional information or mechanism in addition to the testimonies exchanged among agents. For example, Jurca and Faltings propose to build a stand-alone micropayment system which gives witnesses incentives to not give unfair testimonies [11]; ReGreT depends on the social relationship between the witness and the trustee to determine the witness’ credibility [12].

In the proposed credibility model, the truster creates a profile for each witness to maintain the witness’ past testimony-reporting. Each witness’ profile is built as a contingency table of its past testimonies on the trustees and the truster’s own interaction experiences with the corresponding trustees. Then, the truster evaluates the credibility of a witness as a statistic association between the two variables stored in the profile.

Further, the proposed credibility model also guides trusters on how to filter and aggregate testimonies based on the evaluation of credibility: only those testimonies provided by the witnesses with a credibility higher than the truster’s own confidence are aggregated; and those testimonies are aggregated as a weighted average of individual testimonies with the weights derived from the credibilities of the corresponding witnesses.

The rest of this paper is organized as follows. The next section reviews related work. The main components of the proposed model are presented in Section 3 - 5: Section 3 describes the component to build and maintain a witness’ profile; the credibility metric is presented in Section 4; Section 5 discusses how to filter and aggregate the testimonies. The proposed credibility model’s effectiveness is empirically evaluated in Section 6. Finally, Section 7 concludes with a summary of contributions.

2. Besides past behavior, the trustee also has other common agent properties, such as ID and name.

3. The truster’s own confidence in evaluating the trustee’s trustworthiness is calculated in a manner similar to how witness’ credibility is calculated, which will be detailed in Section 5.
and future work.

2 RELATED WORK

The avoidance or mitigation of the adverse influence of unfair testimonies is one of the fundamental issues in the research on computational trust in MAS [7]. Although there are still trust models that ignore the presence of unfair testimonies, there have been recent attempts to address the issue.

In the model proposed by Jurca and Faltings [11], a micropayment system is established to maintain that agents who tell the truth when reporting testimonies will gradually gain money (virtual currency), while those who give unfair testimonies will gradually lose money. In brief, it tackles the presence of unfair testimonies by giving agents incentives to not provide unfair testimonies. The primary disadvantage of the micropayment system is that it needs to be separate from the one used to regulate agents’ ordinary transactions. However, in realistic settings, it is not practical to establish two independent micropayment systems that can be accepted by agents representing different stakeholders.

Whitby et al. proposed a filtering method [13] in the context of Beta Reputation System (BRS) [14]. In BRS, testimonies are given in the form of counts of successful and unsuccessful interactions with the trustees. Each pair of counts models a Beta distribution. All the available testimonies are linearly aggregated to generate the majority opinion, which also models a Beta distribution. An unfair testimony is identified by testing whether it is outside the q quantile and \((1 - q)\) quantile of the majority opinion. If the test is positive, it is considered a possible unfair testimony and will be discarded. Then the majority opinion is updated by aggregating the remaining testimonies before a new round of filtering. This filtering method is time-consuming due to its iterative nature. Moreover, it is tightly coupled with BRS, and requires the testimonies to be given in a form that is specific to BRS, which hinders its wider application.

In ReGreT [12], the adverse impact of unfair testimonies is reduced by aggregating the testimonies as a weighted mean of all the available testimonies, with the weights correspond to the credibilities of the corresponding witnesses. It applies predefined fuzzy rules to determine the credibilities of witnesses, which are based on the social relationship between the witnesses and the trustees, e.g. whether a witness and a trustee are from the same organization. Nevertheless, such relational information is often not available in MAS, which limits its applicability in a wider scale.

In [15], testimonies are also aggregated as a weighted mean of each individual testimony, in which each witness is assigned a weight according to its credibility. Each witness’ credibility is initialized to the maximum possible value, and is updated using a variant of the Weighted Majority Algorithm [16]. If the last testimony given by a witness is similar to the truster’s trust in the trustee after interacting with that trustee, the witness’ credibility remains unchanged. Otherwise, as a penalty, the witness’ credibility is decreased. The penalty is determined according to the distance between the testimony and the truster’s own trust evaluation. This approach does not depend on extra mechanisms or information and can be applied in most trust models. However, with this approach, each agent’s credibility is monotonically decreasing. With such limitation, an agent cannot enhance its credibility no matter how much effort is paid to improve the fairness of its testimonies.

It is thus observed that, existing attempts are not readily useful because they either require additional information which is not always available in realistic settings, e.g. approaches in [11], [12], or are tightly coupled with certain specific trust models, e.g. approach in [13]. Against this background, we propose a credibility model, which works without assuming the availability of additional information in addition to the testimonies exchanged among agents. Further, it is a generic approach that can be applied in most computational trusts model to tackle the presence of unfair testimonies.

3 MAINTAINING THE WITNESS’ PROFILE

The proposed credibility model has three primary components: (1) a component used by the truster to maintain a profile for each witness agent encountered; (2) a metric used by the truster to evaluate the witness agent’s credibility; and (3) a component used by the truster to filter and aggregate the received testimonies. This section discusses the first component, while the other two are discussed in the subsequent sections.

3.1 Notation

For ease of presentation, we define some notations which will be used throughout this paper.

We denote the set of all agents in MAS as \(\mathcal{A} = \{a_1, a_2, \ldots, a_n\}\). Before an agent, i.e. the truster, interacts with another agent, i.e. the trustee, it evaluates the amount of trust it should place in the trustee. The truster is denoted as agent \(a_s\), while the trustee as \(a_o\).

The trust evaluation is an aggregation of agent \(a_s\)’s direct trust in \(a_o\) and third-party testimonies. The direct trust in \(a_o\) is derived based on \(a_s\)’s previous interactions with \(a_o\), and is denoted as \(TP_{s,o}\). The trust evaluation derived from an aggregation of \(TP_{s,o}\) and third-party testimonies is called truster \(a_s\)’s overall trust in \(a_o\), which is denoted as \(TA_{s,o}\).

\(TA_{s,o}\) is the pre-interaction trust evaluation\(^4\). If \(a_s\) decides to interact with \(a_o\) after evaluating \(TA_{s,o}\), it will evaluate \(a_s\)’s trustworthiness again after the interaction with the new observation of \(a_o\)’s behavior in the interaction\(^5\). This is the post-interaction trust evaluation, and is denoted as \(TP'_{s,o}\). It is made only with \(a_s\)’s personal

4. See the attribute trust in Fig. 2.
5. See the attribute outcome in Fig. 2.
interaction experience with \( a_o \) (including the new experience in the interaction). \( TP'_{s,o} \) is obtained mainly to help \( a_s \) update the profiles of those witnesses who have given their testimonies for the evaluation of \( TA_{s,o} \). Information recorded in the witness’ profile is used by \( a_s \) to evaluate the corresponding witness’ credibility. \( a_s \)’s evaluation of witness agent \( a_w \)’s credibility is denoted as \( TT_{s,w} \).

### 3.2 Pre-processing the testimonies

A witness would report its own evaluation of direct trust in the trustee as the testimony to the trustee. Due to the probabilistic nature of trust [17], the testimonies usually take continuous values in the range of \([0,1]\), like those in [11], [12], [13], [15]. The proposed credibility model pre-processes the continuous testimonies to discrete ratings with a scale of \( Z \) before they are stored in the witness’ profile. The pre-processing has certain advantages, which will be discussed in the next subsection. The pre-processing is done by (1) choosing a scale for the discrete ratings (denoted as \( Z \)), (2) stratifying the continuous range \([0,1]\) into \( Z \) bins, and (3) representing the testimonies using the indices of the corresponding bins which contain the testimonies.

### 3.3 Organization of the witness’ profile

After pre-processing, the testimonies are stored in the corresponding witness’ profile\(^7\). The profile \( a_s \) maintains for witness \( a_w \) is organized as shown in Fig. 3. It is basically a contingency table of \( a_w \)’s past testimonies and \( a_s \)’s own post-interaction direct trust in the corresponding trustees.

![Fig. 3: Agent’s profile of giving testimonies](image)

In Fig. 3, \( Z \) denotes the rating scale used to pre-process the testimonies. \( N \) is the number of successful testimony-reporting by witness agent \( a_w \). A testimony-reporting is successful only if both the following conditions are satisfied: (1) the testimony reported by agent \( a_w \) has been used by \( a_s \) in the pre-interaction trust evaluation, and (2) \( a_s \) interacts with that trustee after the pre-interaction trust evaluation.

Each cell \( n_{ij} \) captures the number of successful testimony-reporting in which \( a_w \)’s testimonies are \( j \) while \( a_s \)’s post-interaction evaluations of direct trust in the same trustees are \( i \). And

\[
R_i = \sum_{j=1}^{Z} n_{ij}, \quad C_j = \sum_{i=1}^{Z} n_{ij}, \quad \sum_{i} R_i = \sum_{j} C_j = N \tag{1}
\]

The pre-processing of testimonies discretizes the continuous ratings, which makes the update of the profile easier. That is, when a witness gives a new testimony, its profile can be updated incrementally by changing the count in one of the cells. Moreover, since a witness’ credibility is measured based on the witness’ past testimonies, discrete values are generally more computationally tractable than continuous ones.

### 4 Measuring the Credibility

A credibility metric is proposed to evaluate the witness’ credibility based on the information recorded in the corresponding witness’ profile.

#### 4.1 The Credibility metric

Suppose that, \( a_w \)’s testimonies are the only information available for trustee agent \( a_s \), and \( a_s \) has built a profile of witness agent \( a_w \). Truster agent \( a_s \)’s evaluations of trust in the trustees before the interaction can be considered as an activity to predict the trustees’ trustworthiness after the interactions in two cases [18]:

(1): \( a_s \) makes the evaluations given no testimony from witness \( a_w \).

(2): \( a_s \) makes the evaluations given \( a_w \)’s testimonies.

In Case (1), since the profile contains trustee \( a_s \)’ past post-interaction direct trust, it cannot be worse off by predicting its post-interaction direct trust in the trustee based on its past experiences in interacting with other trustees. More specifically, by referring to the profile, \( a_s \) predicts a particular rating \( i \) with a probability \( R_i/N \). By doing so, the proportion of consistent predictions in the long run is \( \sum_i (R_i/N)^2 \) [18]. Here, a prediction is consistent when trustee \( a_s \)’s pre-interaction evaluation of trust is equal to its post-interaction evaluation of direct trust in the same trustee \( a_w \), i.e. \( TA_{s,a} = TP'_{s,a} \).

In Case (2), in light of witness \( a_w \)’s testimony, the probability that \( a_s \) predicts a particular rating \( i \) is essentially the conditional probability of \( i \)’s occurrence given \( a_w \)’s testimony \( j \)’s presence in \( a_w \)’s profile. More specifically, \( a_s \) predicts \( i \) with a probability as \((R_i)/\sum_i R_i\). Thus, the proportion of the consistent predictions in the long run is \( \sum_i \sum_j (R_i/N)^2 / \sum_i R_i \) [18].

Recall that the credibility of a witness is defined as how much its past testimonies has reduced the truster’s uncertainty regarding the trustees’ behavior. Such a reduction of uncertainty can be quantified as the reduction of inconsistent predictions given the witness’ testimonies, i.e. the reduction of inconsistent predictions.
as truster $a_s$ goes from Case (1) to Case (2). Given this, $TT_{s,w}$ is evaluated as:

$$TT_{s,w} = \frac{(1 - \sum_i (R_i/N)^2) - (1 - \frac{1}{N} \sum_j (n_{ij}/C_j))}{1 - \sum_i (R_i/N)^2}$$

$$= \frac{N \sum_i \sum_j \frac{n_{ij}^2}{C_j} - \sum_i R_i^2}{N^2 - \sum_i R_i^2}$$

(2)

$TT_{s,w}$ takes values in the range of $[0, 1]$. $TT_{s,w} = 0$ implies that witness $a_w$’s testimonies do not reduce $a_s$’s uncertainty at all. $TT_{s,w} = 1$ implies perfect predictability given $a_w$’s testimonies. Note that different trusters may have different evaluations of credibility about the same witness. This is because credibility is a subjective measurement of the usefulness of the witness’ testimonies, and different trusters may have different experiences in receiving testimonies from the same witness.

### 4.2 Bootstrapping the profile and credibility

Intuitively, each cell $n_{ij}$ in $a_w$’s profile should be initialized as 0. However, it is obvious from Eq. (2) that this leads to $TT_{s,w} = 0/0$ after $a_w$’s first successful testimony-reporting. In order to avoid the undefined 0/0, the dummy cell is introduced. That is, every cell in $a_w$’s profile is initialized with a dummy value of $1/Z$ instead of 0. Then all the dummy cells in the $j$th column are cleared by replacing them with 0 after agent $a_w$’s first successful testimony-reporting with a testimony as $j$. After that, the corresponding cell in the $j$th column is updated with real count of successful testimony-reporting.

There is another case leading to the undefined 0/0 even with the introduction of dummy cell. That is, each column in the profile has only one non-zero cell, and all the non-zero cells are in the same row. This represents the fact that $a_s$ always has only one possible prediction in light of $a_w$’s testimonies. This implies $a_s$’s perfect predictability of its post-interaction direct trust given $a_w$’s testimonies. Consequently, $TT_{s,w} = 1$ in this case, as discussed before.

After initializing the profile with the dummy cells, the initial value of $TT_{s,w}$ is 0 according to Eq. (2). However, $TT_{s,w} = 0$ implies that witness $a_w$’s testimonies are not useful at all and should not be aggregated. This consequently makes $TT_{s,w}$ remain as 0. A non-zero initial credibility is hence necessary to bootstrap the testimony aggregation and future credibility update.

To this end, $TT_{s,w}$ is initialized to be close to its updated value after $a_w$’s first successful testimony-reporting to truster $a_s$, where there is only one non-zero cell in $a_w$’s profile. Assume that the only non-zero cell is in the $p$th row and $q$th column, i.e. $n_{pq} = 1$. By applying Eq. (2) on the updated profile, it is obtained that $TT_{s,w} = 1/(Z + 1)$. Given this, truster $a_s$ initializes witness $a_w$’s credibility as $TT_{s,w} = 1/(Z + 2)$, because $1/(Z + 2)$ is close to $1/(Z + 1)$, which helps to smoothen the credibility evolution after $a_w$’s first successful testimony-reporting.

Moreover, $1/(Z + 2) < 1/(Z + 1)$. This corresponds to the intuition that $a_w$ is more credible after its first testimony-reporting to $a_s$ if its testimony has reduced $a_s$’s uncertainty.

### 4.3 Credibility evolution

$TT_{s,w}$ is updated every time $a_w$ makes new successful testimony-reporting to agent $a_s$. Update of $TT_{s,w}$ is triggered each time after truster $a_s$ observes trustee $a_o$’s new behavior in the interaction and obtains its post-interaction direct trust in $a_o$. The update of $TT_{s,w}$ is carried out with the procedure shown in Fig. 4.

**Procedure TT_Update(s,w)**

1. if $a_w$ is a newly-encountered witness for $a_s$, then
2. $a_s$ creates a profile for $a_w$ to record its testimony-reporting
3. for $c = 1$ to $Z$ do
4. for $r = 1$ to $Z$ do
5. $n_{rc} = 1/Z$ {fill in the dummy cell}
6. if $a_s$ interacts with $a_w$ after the pre-interaction trust evaluation then
7. $a_s$ evaluates the post-interaction direct trust in $a_w$ with the new observation of $a_o$’s behavior in the interaction
8. assume the post-interaction direct trust is $i$
9. if $a_w$ makes a successful testimony-reporting with a testimony as $j$ for the first time then
10. for $r = 1$ to $Z$ do
11. $n_{rj} = 0$ {clear the dummy cell in $j$th column}
12. $n_{ij} = n_{ij} + 1$
13. $a_s$ evaluates $TT_{s,w}$ with Eq. (2) based on the updated profile

Fig. 4: Update of witness $a_w$’s profile and credibility

### 5 Testimony filtering and aggregation

Before aggregating the testimonies to obtain the overall trust that it should place in trustee $a_o$, agent $a_s$ evaluates the credibility of all the witnesses whose testimonies on $a_o$ have been received. Then it filters and selects a subset of testimonies to be aggregated.

The testimony filtering is done by comparing the corresponding witness’ credibility against $a_s$’s confidence of its own evaluation of direct trust. In order to evaluate its confidence, truster agent $a_s$ also maintains a profile for itself, which is organized in the same form as Fig. 3. Nevertheless, the cells are associated with different meaning in this case. Each cell $n_{ij}$ in the $a_s$’s own profile records the number of interactions in which $a_s$’s pre-interaction direct trust in the trustees are $j$ while post-interaction ones are $i$. Its confidence, denoted as $TT_{s,s}$, is derived with Eq. (2) based on this profile.

Then only the testimonies contributed by the witnesses that are not less confident than truster $a_s$ itself are aggregated. That is, the testimony reported by witness agent $a_w$ will be selected for aggregation if

$$TT_{s,w} \geq TT_{s,s}$$

where $TT_{s,w}$ is witness $a_w$’s credibility.
A conservative view towards the testimonies is employed when aggregating testimonies. That is, each testimony is assumed to be a possible unfair testimony in the first place, and is adjusted to reduce its adverse effect. The testimony is adjusted based on the corresponding witness’ past testimony-reporting recorded in its profile. Suppose witness \( w \) gives a testimony of \( j \) to trustee \( a_s \). As discussed in Case (2) in Section 4, given witness \( w \)'s testimony \( j \), trustee \( a_s \) can predict its direct trust in trustee \( a_o \) as \( v \) with a probability \( \frac{a_{w,s}}{C_j} = n_{v,j}/C_j \). Put it in another way, trustee \( a_s \) adjusts witness \( w \)'s testimony \( j \) to another rating \( v \) with a probability \( n_{v,j}/C_j \), i.e.:

\[
\Pr(j \rightarrow v) = \frac{n_{v,j}}{C_j}
\]

(4)

The rationale of such adjustment is that: if, in the past, each time witness \( w \) gives a testimony as \( j \), trustee \( a_s \)'s post-interaction direct trust in the trustee is often found to be \( v \). Then, most probably, its direct trust in the trustee would be \( v \) if \( w \) gives a testimony as \( j \) again.

If trustee \( a_s \) accepts a testimony from a newly-encountered witness, this testimony is used directly without adjustment as there is no information recorded in the profile to support the adjustment.

An example of testimony adjustment is given below. Suppose the testimonies are pre-processed to 5-scale discrete ratings. Witness agent \( w \) has made 5 successful testimony-reporting to trust agent \( a_s \):

(i) Agent \( w \) gave three testimonies of 5 on three trustees, and \( a_s \)'s post-interaction direct trust in the corresponding trustees are 3, i.e. \( n_{35} = 3 \).

(ii) Agent \( w \) gave another two testimonies of 4 on two trustees, and \( a_s \)'s post-interaction direct trust in the two corresponding trustees are 3 and 4 respectively.

Witness agent \( a_s \)'s profile after five successful testimony-reporting becomes as Fig. 5 shows.

![Table: Agent \( w \)'s profile after 5 successful testimony-reporting](image)

Now, assume \( w \) gives a new testimony to \( a_s \), and it is 4. Then, it is adjusted as: \( \Pr(4 \rightarrow 3) = n_{34}/C_4 = 1/2 \). \( \Pr(4 \rightarrow 4) = n_{44}/C_4 = 1/2 \). That is, it adjusts this testimony to 3 and 4 with equal probabilities. If \( w \)'s new testimony is 5, then it is adjusted to 3 with a probability as \( n_{53}/C_5 = 3/3 = 1 \).

Finally, the testimonies (after filtering and adjustment), including \( a_s \)'s own pre-interaction direct trust, are aggregated to obtain \( a_s \)'s overall trust in the trustee as a weighted mean of the selected testimonies:

\[
T_{A_{s,o}} = \frac{\sum_{w \in W_{s,o}} \sum_{w \in W_{s,o}} (TT_{s,w} \ast TP'_{w,o})}{\sum_{w \in W_{s,o}} \sum_{w \in W_{s,o}} TT_{s,w}}
\]

(5)

Here \( W_{s,o} \) is a set of witnesses whose testimonies will be aggregated, and \( TP'_{w,o} \) is witness \( w \)'s testimony after adjustment. Each testimony is given a weight as the corresponding witness’ credibility. Truster \( a_s \)'s own pre-interaction direct trust in \( a_o \) is aggregated without adjustment, and is given a weight as its confidence \( TT_{s,a} \).

It is noted that \( T_{A_{s,o}} \) obtained with Eq. (5) takes continuous values in the range of \([0, 1]\). In scenarios where the truster has a list of candidate partners from which it chooses one to interact with, this result can be used directly. This is because usually a rational agent would choose to interact with the one with the highest trustworthiness, and the results derived with Eq. (5) are enough to rank the candidates. However, in scenarios where the truster needs to determine whether to interact with a specific trustee, results obtained with Eq. (5) may need to be mapped to the range of \([0, 1]\), since the decision may be made based on whether the trustee’s trustworthiness is higher than a predefined threshold.

Fig. 6 summarizes the proposed credibility model.

![Fig. 6: The credibility model](image)

6 EMPIRICAL EVALUATION

Conventionally, the effectiveness of methods to tackle the unfair testimonies is measured by how much the methods can keep the trust evaluations’ accuracy in the presence of unfair testimonies [12],[13],[14],[15]. However, agents in MAS usually are concerned not only about the accuracy of trust evaluations, but about whether and how much they can gain by using the methods to choose the interaction partners. Therefore, it is desired that the effectiveness of the proposed model...
can be studied in a more practical manner. To serve this purpose, we have designed a testbed that simulates the interactions among agents as real as possible, and measured the proposed credibility model’s effectiveness with more practical metrics.

6.1 The trust model
As discussed before, credibility is decoupled from trust, and the proposed credibility model can be applied in most existing computational trust models to tackle the presence of unfair testimonies. Nevertheless, to make the studies possible, Beta Reputation System is implemented as the trust model in the testbed. It is chosen because it is widely used in existing work, e.g. [13], [19], which makes it possible to conduct an objective comparison against related work. Moreover, Beta Reputation System uses continuous values to represent the trust evaluation results. Implementing such a model can validate the applicability of the proposed credibility model in models where continuous ratings are used to represent the trust evaluation outcomes. Due to the space constraint, readers are referred to [14] for the details of Beta Reputation System. In the rest of this paper, Beta Reputation System is denoted as BRS for short.

6.2 The testbed
A testbed is designed to simulate a scenario that a number of agents utilize other agents’ services to solve problems. There is cost and gain associated with each service usage. Those service-providing agents have differing behaviors, and some may fail to solve the problem after receiving the payment. BRS is applied by the agents to select the service-providing agents, and the proposed credibility model is applied to mitigate the adverse effect of unfair testimonies in the selection process.

The goal of the study is to investigate the proposed credibility model’s effectiveness in mitigating the adverse effect of unfair testimonies. The effectiveness is measured by how much benefit (in terms of agent’s cost and gain) the proposed credibility model could bring to the agents. BRS and the proposed credibility model are made the only factor that will influence an agent’s cost and gain by excluding all other factors, e.g. negotiation of the payment between the involved agents.

Ideally, the effectiveness should be evaluated in all possible situations. However, the space for all possible situations would be prohibitively large, which renders such an exhaustive investigation impractical. To address this problem, we apply settings that are believed to be typical for some factors. While for some other factors, we introduce randomness to simulate the realistic settings as much as possible. In doing so, the effectiveness can be reasonably evaluated in a wide range of settings.

There are some agents that are assigned a number of problems to solve. There are also agents that provide services that can be utilized to solve the problems. It is assumed that an agent who provides services does not use services provided by others. Thus, the agents in the testbed can be separated into two groups: a set of service-providing agents called service providers, and a set of service-consuming agents called service consumers. And without loss of generality, it is assumed that those provider agents all provide the same service.

Each consumer agent is assigned $N_m$ problems. To solve a problem, each service consumer agent needs to decide whose service to use. It is assumed that each consumer agent is always able to see all the providers available in the testbed. The selection process depends on the consumer agent’s evaluations of the providers’ trustworthiness. As rational agents, consumers will select the provider with the highest trustworthiness.

After a provider is selected, there is an agreement between the consumer and the selected provider, which specifies the provider’s obligations. The outcome of the interaction is binary: successful if the provider fulfills its obligations, or unsuccessful if the provider defaults.

After the interaction, the consumer agent records the interaction outcome to update its experience with the selected provider. If the outcome is successful, the consumer agent proceeds to solve the next problem. Otherwise, the consumer will keep trying other providers according to the rank of their trustworthiness until it finds one to solve the problem. In the worst case, the consumer fails to solve the problem after finding all the providers have defaulted their obligations.

The behavior of a provider is controlled by a willingness factor, which is defined as the probability that it would fulfill its obligations. The higher the “willingness”, the more likely the provider will fulfill its obligations, and vice versa. Providers with various levels of willingness are introduced into the testbed: excellent providers, good providers, ordinary providers, and bad providers. The willingness of the four different types of providers are given in Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>willingness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>0.8</td>
</tr>
<tr>
<td>Good</td>
<td>0.6</td>
</tr>
<tr>
<td>Ordinary</td>
<td>0.4</td>
</tr>
<tr>
<td>Bad</td>
<td>0.1</td>
</tr>
</tbody>
</table>

TABLE 1: Willingness of providers

It is usually the case that provider’s behavior is not static. Instead, their behaviors change from one interaction to another. To simulate such dynamism, we also implement the fluctuation of provider’s willingness: a provider’s initial willingness is set according to the provider’s nature as specified in Table 1; then it can change its willingness in each subsequent interaction. Three strategies of fluctuation are simulated, namely, increasing, decreasing from or remaining the same as

9. Since the goal is to investigate the proposed credibility model’s effectiveness but not how the problems are solved, it is assumed that consumer’s problem is solved if and only if the selected provider fulfills its obligations.
that in the previous interaction. The magnitude of each change is randomly chosen from the range of [0, 0.01]. Each provider has an equal probability to choose one of the three fluctuation strategies. That is, the probability (denoted as $p_{f\text{lux}}$) that each provider chooses one of the three strategies in each interaction is set to 1/3 identically. The willingness of provider is always maintained within the range of [0, 1].

The four types of providers have different population in the testbed. It is impractical to explore completely all the possible configurations of provider population. Instead, we choose a population setting as shown in Table 2. Such a population can be considered hostile to consumer agents as only a small percentage of providers are highly willing to fulfill their obligations, most of the providers only do it occasionally, and a large percentage of providers tend to default. Such a population setting is chosen due to the consideration that if the proposed credibility model is able to show its effectiveness in such a hostile environment, its effectiveness can generally be guaranteed in other settings which are relatively more friendly to the consumers.

<table>
<thead>
<tr>
<th>Category</th>
<th>percentage among all the providers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>10%</td>
</tr>
<tr>
<td>Good</td>
<td>10%</td>
</tr>
<tr>
<td>Ordinary</td>
<td>40%</td>
</tr>
<tr>
<td>Bad</td>
<td>40%</td>
</tr>
</tbody>
</table>

**TABLE 2: Population of the providers**

For a consumer agent, say $a_s$, there is a cost associated with each interaction with provider $a_o$, which is calculated as $C \times T A_{s,o}$. $T A_{s,o}$ is agent $a_s$’s pre-interaction evaluation of trust that it should place in provider $a_o$. $C$ is a constant which denotes the cost if the selected provider $a_o$ is evaluated to have full trustworthiness (i.e. $T A_{s,o} = 1$). The gain consumer $a_s$ would obtain each time a problem is solved by the selected provider is calculated as $G \times L_o$. $L_o$ denotes the willingness of the selected provider when it solves the problem. $G$ is a constant which denotes the gain if the problem is solved by a provider with full willingness (i.e. $L_o = 1$). By doing so, each consumer agent can gain more if it manages to find a provider with higher willingness to solve the problem with fewer tries. Its gain decreases with more tries. The consumer may even lose its profit if it tries too many untrustworthy providers before the problem is solved.

### 6.3 Presence of unfair testimonies

As discussed earlier, unfair testimonies can happen because of the differing views of the witness and the trustee. This is implemented as individual unfair testimonies in the testbed.

**Individual unfair testimonies** mean that providers do not collude with consumers, and each consumer gives unfair testimonies individually. We simulate five categories of consumer agents with different strategies in giving testimonies: $H, P, HP, N,$ and $HN$. Consumer agents of category $H$ always reveal honestly their own evaluations of direct trust in the providers as testimonies. Agents in category $P$ and $HP$ give ballot-stuffing unfair testimonies, i.e. unfairly positive testimonies that are higher than their real evaluations of direct trust in the providers; while agents in category $N$ and $HN$ give badmouthing unfair testimonies, i.e. unfairly negative testimonies that are lower than their real evaluations of direct trust in the providers. The unfair testimonies are generated by applying offset to the witnesses’ real evaluations of direct trust in the provider. The offset is positive for ballot-stuffing unfair testimonies, and negative for badmouthing unfair testimonies. The magnitude of offset for category $P$ and $HP$, or $N$ and $HN$ is randomly chosen from the range of $(0.1, 0.4)$, i.e. the agents of these two categories are giving moderately unfair testimonies. The range for category $HP$ and $HN$ is $(0.8, 1)$, i.e. the agents of these two categories are giving highly unfair testimonies.

There are various configurations of unfair testimonies presented in the testbed. Each configuration contains a mix of consumers of the following category combinations: \{H, P, HP\}, or \{H, N, HN\}. In each configuration, the agents of category $P$ and $HP$, or $N$ and $HN$ have equal presence. Configurations differentiate from each other in the ratios of consumer agents giving unfair testimonies. For the convenience of reference, each configuration is named with the nature of unfair testimony (i.e. Pos for ballot-stuffing and Neg for badmouthing) and the ratio of agents giving unfair testimonies. For example, Neg60 denotes the configuration in which 30% of the consumers are of category $N$, 30% are of the category $HN$, and the rest 40% are of category $H$. There is a special configuration which is named Hon, where all the consumer agents are of category $H$.

Unfair testimonies may also be caused by some agents deliberately give unfair testimonies on some trustees. In this case, the unfair testimonies are usually given with a clear target to manipulate the trustee agents’ evaluations of trust in those target trustee agents. This is implemented as another type of unfair testimonies, i.e. the collusive unfair testimonies. Collusive unfair testimonies mean that a number of providers collude with consumers, who intentionally give testimonies in favor of the colluding providers in order to promote unfairly the colluding providers’ trustworthiness in the eyes of the non-colluding consumers.

To simulate the collusive unfair testimonies, the Bad providers (see Table 1 for the definition of Bad provider) are chosen as colluders. Since the Bad providers generally have the lowest evaluations of trust by the consumer agents if the evaluations are not affected by the presence of unfair testimonies, they can only attract a few consumers to use their services. Hence, they have the motivation to collude. The configuration of collusive unfair testimonies is similar to the case of individual unfair testimonies. Nevertheless, the testimonies given
by the colluding consumers are always ballot-stuffing. Hence, the agents giving collusive unfair testimonies are only from the the category combination of \{\text{H, P, HP}\}.

### 6.4 Methodology

The consumer agents in the testbed are separated into three groups. Each group consists of (1) agents equipped with the proposed credibility model, (2) agents with no mechanism to tackle the presence of unfair testimonies, and (3) those with other methods to tackle the presence of unfair testimonies. By doing so, an objective comparison with related work is made possible since they operate in identical settings.

For the third group, the method proposed by Yu and Singh\(^{10}\) is chosen. It is selected because: first, its applicability has already been validated by some successful applications, such as P2P system in [20]; second, other than this, most of the notable existing methods requires additional information which is not available in most of the cases, such as those in [11], [12].

Other than the three groups mentioned above, there is another group of consumers, who do not aggregate the third-party testimonies at all. This group is denoted as NoAggr, and it is introduced as the benchmark to study whether the aggregation of testimonies improves the consumers’ performance in conducting trust evaluations. Group (1)-(3) are labeled as Cred, NoCred, and YS respectively. All the four groups are of equal size, and the number of agents in each group is denoted as \(N_c\).

The effectiveness of the credibility model is measured in terms of the benefit agents achieve by using it. In the testbed, BRS is applied in all the four groups to help agents identify the right service provider. The only difference between the different groups of agents is that each employs different strategies to combat the unfair testimonies. Additionally, the testbed rules out other factors that can potentially influence an agent’s profit, e.g. negotiation between agents regarding the payment. Therefore, the profit gained by agents in different groups is a good proxy of the effectiveness of the corresponding strategies in addressing the unfair testimonies. Given this, we record the gain of each agent in solving each problem. The average over all agents’ gains in each individual group is calculated and taken as each group’s performance. This measurement metric is termed as gain. The higher the gain, the more effective the corresponding strategy is.

Besides this gain metric, we also measure the collusion power when evaluating different strategies’ effectiveness in the presence of collusive unfair testimonies. The collusion power is defined as follows:

\[
\text{collusion power} = \frac{\sum_{a_s \in A_{nc}} \#try(a_s)}{|A_{nc}| \times N_m}
\]

\(N_m\) is the number of problems each consumer is assigned to solve. \(A_{nc}\) denotes the set of non-colluding consumers, \(|A_{nc}|\) is the number of consumers in the set of \(A_{nc}\), \(a_s\) is a consumer in this set, and \(#try(a_s)\) denotes the number of times consumer \(a_s\) has tried to invoke any colluding Bad provider.

If the adverse effect of collusive unfair testimonies are not fairly addressed, the collusion is able to manipulate the non-colluding consumers’ evaluations of trust in the colluding providers, and attracts more non-colluding consumers to use their services, which leads to a higher collusion power. Hence, a low value of collusion power is expected to show the effectiveness. For example, suppose there are 20 non-colluding consumers (i.e. \(|A_{nc}| = 20\), and each of them needs to solve \(N_m = 100\) problems. If these 20 non-colluding consumers have attempted to invoke the services provided by the colluding providers 4000 times, the collusion power in this case would be \(\frac{4000}{20 \times 100} = 2\). This value can be interpreted that each of the non-colluding consumers has tried the colluding Bad providers twice on average for every problem.

### 6.5 Experimental setup

In summary, the running of the testbed is controlled by the following parameters:

- The agent population. In each experiment, the testbed is populated with \(N_p = 10\) providers. There are four groups of consumers. Each group contains \(N_c = 100\) consumers.
- The number of problems each consumer needs to solve is denoted as \(N_m\). It is set as 200.
- The initial balance of each consumer agent, i.e. \(G_{init}\), is set as 50.
- The constant \(C\) to calculate the cost consumer needs to pay for each service usage is set as 1.
- The constant \(G\) to calculate the gain of consumer if a problem is solved is set as \(G = 5\).
- The providers’ willingness. It is set according to Table 1 and Table 2.
- The probability that each provider changes its willingness in each interaction, i.e. \(p_{fluc}\). It is set as \(p_{fluc} = 1/3\).

BRS is utilized by the agents to evaluate providers. BRS is mainly controlled by the decay factor (denoted as \(\lambda\)) and the number of ratings that are taken into account when generating the rating summary (denoted as \(W\)). These two parameters are set as \(\lambda = 0.9\) and \(W = 10\).

Table 3 summarizes the parameters.

There are also a number of variables, which control the presence of unfair testimonies in the testbed, such as the configuration of unfair testimonies. We apply different settings of these variables to generate various configurations of unfair testimonies. Five individual experiments are run with each setting of these variables, and the average of the aforementioned metrics obtained in the five experiments are taken as the final result of the corresponding configuration.

\(^{10}\) Due to the space constraint, readers are referred to [15] for details about this method.
The proposed credibility model itself also has a parameter for fine tuning, which is the scale of discrete ratings that represent the discretized testimonies, namely $Z$ in Fig. 3. The default value of $Z$ is set as 10 unless otherwise specified. We also investigate the impact of its value over the effectiveness of the proposed credibility model by running the experiments with different values of $Z$.

### 6.6 Effectiveness in the presence of individual unfair testimonies

First of all, as the benchmark, we study each group’s performance with the unfair testimonies configuration $\text{Pos100}$, where no agent gives unfair testimonies. Table 4 summarizes the average gain of each group. It is observed that group $\text{NoAggr}$ achieves the lowest performance (in terms of gain). Fig. 7 plots the average gain of agents in each individual group for each problem. It is observed from Fig. 7 that for almost every problem, group $\text{NoAggr}$ is outperformed by other groups. This observation shows that solely depending on an agent’s private interaction experiences with the provider is not sufficient to make a trust evaluation that truly reflects the provider’s behavior. This observation also confirms the effect of third-party testimonies [10], that aggregation of third-party testimonies does reduce the consumer agent’s uncertainty about the provider’s behavior.

We then run experiments with different configurations of unfair testimonies to see whether each group is able to maintain its performance as in the case where there were no unfair testimonies.

Each group’s performance is first studied in the presence of unfair testimonies $\text{Pos100}$. Table 5 lists each group’s average gain in this case, while Fig. 8 plots the average gain of the agents in each individual group for each problem. In the presence of $\text{Pos100}$ unfair testimonies, group $\text{NoAggr}$ manages to maintain its

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_p$</td>
<td>number of providers</td>
<td>10</td>
</tr>
<tr>
<td>$N_c$</td>
<td>number of consumers in each group</td>
<td>100</td>
</tr>
<tr>
<td>$N_m$</td>
<td>number of problems assigned to each consumer</td>
<td>200</td>
</tr>
<tr>
<td>$G_{\text{init}}$</td>
<td>initial balance of consumer agent</td>
<td>50</td>
</tr>
<tr>
<td>$C$</td>
<td>The constant used to calculate consumer agent’s cost for each service usage</td>
<td>1</td>
</tr>
<tr>
<td>$G$</td>
<td>The constant used to calculate consumer agent’s gain for each solved problem</td>
<td>5</td>
</tr>
<tr>
<td>$P_{\text{f Luc}}$</td>
<td>The probability that each provider changes willingness</td>
<td>1/3</td>
</tr>
<tr>
<td>$A$</td>
<td>decaying factor</td>
<td>0.9</td>
</tr>
<tr>
<td>$W$</td>
<td>number of past ratings taken into account</td>
<td>10</td>
</tr>
</tbody>
</table>

TABLE 3: Experimental Parameters

<table>
<thead>
<tr>
<th>Group</th>
<th>Average Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cred</td>
<td>3.45</td>
</tr>
<tr>
<td>NoAggr</td>
<td>2.66</td>
</tr>
<tr>
<td>NoCred</td>
<td>3.09</td>
</tr>
<tr>
<td>YS</td>
<td>2.95</td>
</tr>
</tbody>
</table>

TABLE 4: Performance without unfair testimonies

Fig. 7: Average gain of each group without unfair testimonies

Fig. 8: Average gain of each group in the presence of individual unfair testimonies $\text{Pos100}$
performance as in the case of Hon. This is because it is not influenced by the presence of unfair testimonies since it does not aggregate testimonies at all. However, group NoCred’s performance has been greatly aggravated by the unfair testimonies, and it is outperformed by the other groups in almost every problem as shown by Fig. 8. This is due to the fact that it blindly aggregates all the testimonies, including the unfair testimonies. Compared with group NoCred, group YS mitigates the adverse influence of unfair testimonies, but its performance is not as good as in the case of Hon. It is also outperformed by group Cred, which is able to maintain its performance at a comparable level as in the case of Hon.

<table>
<thead>
<tr>
<th>Group</th>
<th>Average Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cred</td>
<td>3.45</td>
</tr>
<tr>
<td>NoAggr</td>
<td>2.64</td>
</tr>
<tr>
<td>NoCred</td>
<td>2.37</td>
</tr>
<tr>
<td>YS</td>
<td>2.71</td>
</tr>
</tbody>
</table>

TABLE 5: Performance in the presence of individual unfair testimonies Pos100

Experiments have been also conducted to study the four groups’ performance in presence of other configurations of unfair testimonies, namely, Pos80, Pos60, Pos40, Pos20, Neg20, Neg40, Neg60, Neg80, and Neg100. Due to the space constraint here, instead of showing the detail of each individual experiment, an overview of the average gains of agents in the four different groups for each problem are presented in Fig. 9. Group NoAggr’s performance is relatively stable for all configurations of unfair testimonies studied. As discussed before, this is because it is not influenced by the presence of unfair testimonies as it does not aggregate testimonies. However, since the provider’s behavior is changing from one interaction to another, solely depending on the consumer agent’s own interaction experience with the provider is not sufficient to make a trust evaluation that truly reflects the provider’s behavior. Hence, group NoAggr achieves a performance lower than the other groups that aggregate testimonies (i.e. group Cred and group YS, but not NoCred), though its performance is stable in the presence of unfair testimonies.

Group NoCred’s performance exhibits a great fluctuation in the presence of various configurations of unfair testimonies. Generally, group NoCred’s performance (in terms of gain) decreases with more consumers giving unfair testimonies. This is because the presence of unfair testimonies exerts a great impact on the consumer agents’ evaluations of trust in the providers, which confuse the consumer agents’ selection of the right providers (i.e. those with higher willingness to fulfill their obligations). Consequently, the agents in group NoCred need more tries before each problem can be solved, leading to their lower gain for each problem. This observation also exemplifies the adverse effect of unfair testimonies.

Both group YS and group Cred reduce the adverse effect of unfair testimonies, and thus their performance is more stable, as compared to group NoCred. Nevertheless, group YS’s performance is still inferior to that of group Cred. Group YS is even outperformed by group NoCred in the case of Hon. This is because, group YS applies a strategy that is too sensitive to the presence of unfair testimonies. That is, every testimony that is not exactly the same as the trustor’s post-interaction direct trust in the trustee will be considered unfair, and the witness’ credibility will be decreased accordingly as a penalty. However, as there is no agent giving unfair testimonies in the case of Hon, such a sensitive strategy would reduce the fair testimonies’ contribution to the trust evaluations, making it even less effective than the strategy of not measuring the credibility (i.e. as group NoCred does).

As expected, group Cred consistently outperforms group NoAggr and NoCred. Moreover, group Cred also outperforms group YS. Several reasons contributes to group Cred’s superiority over group YS:

1. A consumer agent in group YS first aggregates the testimonies as a weighted mean of all the available testimonies. Then the aggregation of testimonies is linearly combined with the consumer’s own evaluation of direct trust to obtain the overall trust in the provider. The factor assigned to the aggregation of testimonies is based on the number of past interactions with the provider. That is, if the number of interactions the consumer and the provider is not less than a predefined threshold, the testimonies are not aggregated. However, since the behavior of provider is changing from one interaction to another, solely depending on the consumer’s own interaction with the provider and not aggregating the third-party testimonies after a number of interactions with the providers will deteriorate the trust evaluation.

2. In contrast, consumer agents in group Cred apply a more adaptive strategy in aggregating the testimonies. Only those less useful than its own evaluation of direct trust in the provider in each trust evaluation will not be aggregated.

3. Consumers in group YS aggregate all the testimonies without filtering. Although different testimonies are assigned different weights, unfair testimonies still exert certain influence on the consumer agents’ trust evaluation when they are aggregated.
responding witnesses’ credibility, consumer agents in group Cred also adjust each testimony based on the corresponding witness’ past testimony-reporting, which makes a testimony useful even if it may be unfair. With testimony adjustment, the agents are still able to make accurate trust evaluations even when large percentage of agents are giving unfair testimonies.

6.7 Effectiveness in the presence of collusive unfair testimonies

We also study the performance of different groups in the presence of collusive unfair testimonies. As the colluding consumers are always giving ballot-stuffing unfair testimonies, the configurations of unfair testimonies considered are Hon, Pos20, Pos40, Pos60, Pos80. The gain metric is applied again to measure the effectiveness. However, it is only calculated on the non-colluding consumers since the colluding consumers’ unfair testimonies only affect the non-colluding consumers’ evaluations of trust in the colluding Bad providers. Besides the gain metric, we also measure the collusion power. Fig. 10 presents the average gain and collusion power of each group in the presence of various configurations of collusive unfair testimonies. Since the performance of group NoAggr is not influenced by the presence of unfair testimonies, it is not included here.

The first observation from Fig. 10 is that the collusive unfair testimonies do cheat the non-colluding consumers to try the colluding providers more frequently. In Fig. 9, group NoCred is still able to achieve an average gain above 2.3 even in the presence of most powerful individual unfair testimonies, i.e. Pos100 and Neg100. However, as shown in Fig. 10(a), presence of collusive unfair testimonies Pos20 already deteriorates group NoCred’s average gain to as low as 2.3. This observation shows that the adverse effect of collusive unfair testimonies is more powerful than that of individual unfair testimonies.

It is also observed that the proposed credibility model does mitigate the adverse effect of collusive unfair testimonies, as the collusion power for group Cred is kept below 0.25 in the presence of various configurations of collusive unfair testimonies. However, the average gain of group Cred is still lower than that in the presence of individual unfair testimonies. This shows again that the collusive unfair testimonies are more powerful in cheating group Cred than the individual unfair testimonies. This is because, the colluding consumers do not give unfair testimonies on non-colluding providers. Their testimonies on those providers are still considered as useful for the non-colluding consumers. Hence, the colluding consumers are able to achieve a high credibility and then give unfair testimonies on the colluding providers. This makes the non-colluding consumers’ filtering and adjustment of testimonies inaccurate. Nevertheless, due to the reasons discussed in Section 6.6, group Cred still outperforms group YS.

6.8 Influence of Z value

To study the influence of parameter Z on the proposed credibility model’s performance, we applied different value of Z and re-run the experiments. The smallest possible value of Z is 2, that is, the continuous testimonies are pre-processed to become binary ratings. We have also studied the other possible values of Z, i.e. Z = 5, Z = 7 and Z = 10. It is noted that the experiments are re-run with only the presence of collusive unfair testimonies due to the consideration that the collusive unfair testimonies are more powerful in cheating the credibility model, as shown by the results in Section 6.6 and 6.7.

Fig. 11 presents the collusion power obtained with different Z values. It is observed that the credibility model’s effectiveness (in terms of collusion power) is generally improved with the increase of Z value. This is because parameter Z basically controls the credibility model’s sensitivity to the witness’ unfair testimonies. With a smaller value of Z, the discretization of testimonies is coarser, which makes the discrimination between fair and unfair testimonies relatively coarser too.

\footnote{Pos100 is not considered since there is no non-colluding consumer in this case.}
Hence, a witness is able to achieve a high evaluation of credibility even if its testimonies are largely different from the truster’s own post-interaction direct trust. This will consequently decrease the effectiveness of testimony filtering and adjustment. On the other hand, with a larger value of $Z$, the discrimination between honest and unfair testimonies is relatively finer. The evaluation of the witness’ credibility can thus reflect the usefulness of the witness’ past testimonies more truly, which consequently enhances the effectiveness of testimony filtering and adjustment.

There is a trade-off in the selection of $Z$ value. It is expected that the proposed credibility model’s effectiveness would be enhanced with a larger value of $Z$. Nevertheless, the value of $Z$ also determines the size of the profile that the truster maintains for each witness. The storage for maintaining the profiles increases with $Z$. Based on the results shown in Fig. 11, an good choice of $Z$ is in the range of $[5, 10]$, which achieves a balance of effectiveness and storage overhead.

7 CONCLUSION AND FUTURE WORK

In this paper, a credibility model is proposed to mitigate the adverse effects of unfair testimonies. It measures the usefulness of testimony based on how much past testimonies from the same witness have reduced the truster’s uncertainty in evaluating trusters’ trustworthiness. It can be applied in most existing trust models as long as the trust evaluation outcomes are discrete or can be discretized. Experimental results show that the proposed credibility model is more effective than related work in mitigating the adverse effect of unfair testimonies.

The proposed credibility model can be applied to build robust trust systems in any system that can be modeled as multi-agent systems (MAS). Some typical application domains include e-commerce systems [21], Peer-to-Peer content-sharing networks [22], Grid [23], and massive multi-player online role playing games (MMORPGs). It can also be used to enhance the performance of collaborative filtering [24].

One issue left unaddressed in this paper is how the agent is able to find relevant witnesses and collect available testimonies efficiently [25]. We have implemented a word-of-mouth process [15], [20] to facilitate the testimony discovery in the experiments. That is, each agent maintains direct relationships with a number of credible agents, and it requests each of those credible agents to collect testimonies recursively from other agents that the latter believe to be credible. However, this introduces the presence of malicious agents who do not give unfair testimonies directly, but redirect the trusters to other unfair-testimony-giving agents. We have begun to investigate the influence of such malicious agents and how to mitigate their influence.

REFERENCES


Dr. Jianshu Weng is currently a post-doctoral research fellow in the School of Information Systems, Singapore Management University. He received his Ph.D degree from the School of Computer Engineering, Nanyang Technological University, Singapore in 2008. His current research interests include trust modeling in multi-agent systems and influence diffusion in online social communities.

Dr. Zhiqi Shen is currently a Visiting Assistant Professor at Information Engineering Division, School of EEE, NTU. He enjoyed working for industry (Siemens Singapore, MDSI Canada), research institute (i2R/NUS) and university (NTU). His research interests include intelligent agents, computational intelligence, e-learning, mixed reality, semantic web/grid and interactive storytelling/games. He has extensive experiences in transfer the research results into commercialized technologies. He successfully led a number of well know research projects in Singapore and Canada and transferred the research outcome into commercial applications.

Dr. Chunyan Miao is an Assistant Professor in the School of Computer Engineering, Nanyang Technological University (NTU). Her research areas include Agent, Multi-Agent Systems (MAS), Agent Oriented Software Engineering (AOSE), AgentWeb/Grid/Cloud and agents for wireless communications. Her current research works focus on infusing intelligent agents into interactive new media (virtual, mixed, mobile and pervasive media) to create novel experiences and dimensions in virtual world, game design, interactive narrative and other real world agent systems.

Dr. Angela Goh Eek Soong is Professor and Acting Chair in the School of Computer Engineering, Nanyang Technological University. She received her Ph.D from the University of Manchester. Her current research interests include service-oriented computing and semantic web. She is on the editorial board of several international journals, a fellow of the British Computer Society and chartered IT professional, a fellow of the Singapore Computer Society and a member of the Association for Computing Machinery.